Sorghum bicolor – an important species for comparative grass genomics and a source of beneficial genes for agriculture

John E Mullet*, Robert R Klein† and Patricia E Klein*

A high-resolution genetic, physical, and cytological map of the sorghum genome is being assembled using AFLP DNA marker technology, six-dimensional pooling of BAC libraries, cDNA mapping technology, and cytogenetic analysis. Recent advances in sorghum comparative genomics and gene-transfer technology are accelerating the discovery and utilization of valuable sorghum genes and alleles.

Addresses
*Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, Texas 77843, USA
†USDA-ARS, Southern Plains Agricultural Research Center, College Station, Texas 77845, USA

Current Opinion in Plant Biology 2001, 5:118–121

1369-5266/01/$ – see front matter

Published online 20 December 2001

Abbreviations

AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridization
RFLP	restriction fragment length polymorphism

Introduction

Sorghum, a C4 grass that diverged from maize just 15 million years ago, is the fifth most important cereal grown worldwide [1]. This grain and forage crop is especially important in the semiarid tropics because of its unusual tolerance of hot and dry environments. Recently, sorghum has been identified as a key plant species for the comparative analysis of grass genomes, and as a source of beneficial genes for agriculture. Sorghum's relatively small genome (750 million base pairs [Mbp]) [2], extraordinary diversity of germplasm [3,4], and incremental divergence from maize and rice [5], make it ideally suited to aid the discovery and analysis of grass genes through comparative genomics. Here, we review recent advances in sorghum genome technology and information about the sorghum genome that indicate that studies of this species will play a key role in expanding our knowledge of the functions of grass genes. These advances also provide insights into the evolution and architecture of plant genomes.

Accelerated progress on a comprehensive sorghum genome map

High-resolution genome maps are especially important for species such as sorghum that are not currently the target of complete-genome-sequencing efforts. Genome maps enable efficient map-based isolation of genes, targeted genome sequencing, detailed investigation of genome architecture, useful comparison with the genomes of other plants, and association studies that link DNA markers (and genes) to important phenotypes. We have made rapid progress in the construction of an integrated sorghum genome map. This progress has been achieved using a combination of high-throughput amplified fragment length polymorphism (AFLP) DNA marker technology [6••,7], six-dimensional pooling of bacterial artificial chromosome (BAC) clones [6••], cDNA capture technology [8••], and BAC-based fluorescence *in situ* hybridization (FISH) [9] (http://sorghome.tamu.edu).

AFLP DNA marker technology and a recombinant inbred population were used to construct a high-resolution genetic map for sorghum [7]. The 1713 cM map included 2926 loci distributed within ten linkage groups; 2454 of the mapped loci were AFLP products, 136 were simple-sequence repeats (SSRs) [6••], and 203 were restriction fragment length polymorphism (RFLP) loci mapped earlier in sorghum using cDNA or genomic probes [10]. Of the nearly 3000 markers, 692 comprised a LOD ≥ 3.0 framework genetic map with an average spacing of approximately 1 Mbp. The RFLP loci served to align the new high-resolution AFLP-marker-based map with other sorghum and grass genetic maps [10]. The sorghum genetic map, gel images, and primer combinations used to generate polymorphic amplified DNAs for mapping are available on a public website (http://SorghumGenome.tamu.edu).

The construction of the sorghum physical map initially relied on DNA-fingerprinting methods to generate contigs of overlapping BACs [6••]. AFLP technology was then used to join BAC contigs together, and to locate the contigs on the high-resolution AFLP-based genetic map described above [6••]. (AFLP analysis was streamlined by pooling about 25 000 BACs representing approximately 4X genome equivalents in six different ways [rows, columns, plates, and three diagonals] to create 184 pools of DNA [6••]. The distribution of AFLP, SSR, or sequence-tagged site [STS] markers among the 184 pools of DNA allowed BACs containing the DNA markers to be readily identified.) Moreover, because the genetic and physical maps are based on and linked through the same set of AFLP DNA markers, sorghum trait loci that have been mapped using AFLPs can be readily located on the sorghum genome map. This provides direct access to BAC clones containing the genes of interest. In fact, there are two working maps for sorghum; the second genetic and physical map is based on BAC fingerprinting and various hybridization techniques [11,12].

In parallel with the building of genome maps, the overall architecture of the sorghum genome is being characterized using FISH technology with improved resolution and

sensitivity [9]. Key BAC clones from the sorghum genome map, and probes corresponding to centromeres, telomeres, and other chromosomal features, are being located on the sorghum cytogenetic map using FISH. A major discovery from this work is that recombination frequency varies substantially across sorghum chromosome 1, and is especially low in a large region near the centromere (MN Islam-Faridi *et al*., abstract P141, Plant and Animal Genome IX, San Diego, January 2001). FISH analysis has also identified a large block of pericentromeric heterochromatin in chromosome 1. Targeted genome sequencing studies indicate that gene density in sorghum euchromatin is similar to that in rice euchromatin, even though the rice genome (400 Mbp) is significantly smaller than the sorghum genome (750 Mbp). This suggests that the difference in the sizes of the rice and sorghum genomes may reflect the greater amounts of repetitive DNA in sorghum's pericentromeric heterochromatin.

Sorghum gene mapping and the use of expressed gene sequences

Over the past three years, a sorghum expressed sequence tag (EST) project at the University of Georgia has obtained information from 106 296 sorghum ESTs by sequencing both ends of approximately half that many cDNA clones. Clustering of 52 436 3′ ESTs revealed the presence of approximately 15 500 unigenes (MM Cordonnier-Pratt *et al*., personal communication; http://www.botany.uga.edu/~prattlab/). In addition, sorghum ESTs are being located on the sorghum genome map using a novel type of direct selection technology [8••]. In this approach, BAC clones obtained from the sorghum genome map are covalently attached to epi-tubes to facilitate high throughput capture of cDNA encoded by each BAC. A complex mixture of cDNA is then hybridized to the BAC DNA in each tube followed by washing, PCR amplification, and analysis of selected cDNA by sequencing. Two rounds of direct selection are sufficient to normalize the population of captured cDNA sequences, allowing up to 75% of the genes encoded in any region of the sorghum genome to be identified [8••].

This cDNA mapping technology provides an efficient method with which to search for genes in trait loci, and with which to map gene sequences and assess gene density across the sorghum genome. Furthermore, gene sequence information obtained at regular intervals across the sorghum genome will allow the sequence-based alignment of the sorghum genome map with the rice genome sequence, which is rapidly nearing completion (http://rgp.dna.affrc.go.jp/). In parallel, sorghum ESTs are being deployed on microarrays that will be used to characterize global changes in sorghum gene expression that occur in response to abiotic and biotic factors (LH Pratt *et al*., unpublished data). This approach will allow researchers to begin to functionally annotate mapped sorghum genes.

Utility of comparative genome maps, gene sequences and function

Bennetzen [13••] recently reviewed the considerable evidence for significant conservation of gene order in grass genomes and pointed out the potential value of local synteny for cross-species gene discovery. Unfortunately, early reports of the conservation of gene order between sorghum and *A. thaliana* [14] have not proven practically useful for predicting gene order in sorghum and other grasses [15,16]. Detailed sequence analyses of selected homologous regions of several grass genomes have confirmed the significant conservation of gene order within the grasses, but also revealed inversions, duplications, chromosome translocation, and individual gene transfer events [13••,17,18]. Given that the grasses that were compared have diverged over 60 million years and contain different numbers of chromosomes, an important finding is that so few rearrangements were detected.

Comparisons of orthologous gene sequences from related grass species [19] and other flowering plants [20] have demonstrated the utility of this approach for the identification of exons, intron/exon boundaries, and other conserved elements of genes. During the period since the separation of sorghum from maize and rice (approximately 15–20 million years and about 50 million years, respectively [5]), extensive divergence of repetitive sequences has occurred. Nevertheless, the functional portions of genes from these three crops retain sufficient conservation to be readily recognized ([18]; DT Morishige, JE Mullet, unpublished data). Therefore, the acquisition of gene sequences from sorghum will be extremely valuable for the identification and annotation of sequences from the genomes of rice, maize and other grasses.

The analysis of orthologous gene function in a diverse set of species is providing a deeper understanding of the full range of possible plant gene activities. The study of phytochrome B (PHYB), one of several red-light photoreceptors [21], in sorghum and *A. thaliana* provides a good example of the utility of this approach. In sorghum, PHYB is encoded by *Ma3*, one of six loci that regulate flowering time in this species [22]. Mutation of PHYB causes early flowering in both sorghum, a long-day plant, and *A. thaliana*, a short-day plant. PHYB deficiencies in both sorghum and *A. thaliana* cause a constitutive shade-avoidance phenotype. PHYB-deficient sorghum plants synthesize ten times the normal amount of ethylene in cycles with a circadian rhythm [23,24]. Thus, these studies show that PHYB is involved in both shade avoidance and the regulation of flowering time in both a dicot and a monocot, and also revealed novel aspects of PHYB function in sorghum.

Advances in sorghum gene transfer technology

In common with most grasses, sorghum is recalcitrant to transformation. However, Zhao *et al*. [25•] recently reported *Agrobacterium*-mediated sorghum transformation at a frequency of 2%, with the distinct potential for even higher

efficiencies. In addition, Emani *et al.* [26] showed that methylation-based gene silencing is one of the major obstacles to the expression of foreign genes in sorghum. Improvement in sorghum transformation efficiency, together with progress in transposon tagging [27] and virus-induced gene silencing [28], will dramatically enhance the rate and efficacy of the discovery and analysis of sorghum genes.

Successful transfer of useful genes from sorghum to other plants

Sorghum germplasm contains unique genes and beneficial alleles that enable this plant to survive in the arid African environment in which it originated [1]. Sorghum genes providing adaptation to adverse environments or that control other beneficial traits could be usefully introduced into other crops if they could be identified, transferred, and expressed successfully. For example, two sorghum genes, CYP79A1 and CYP71E1, that encode microsomal cytochrome P450s that are involved in the synthesis of cyanogenic glucosides have been transferred to *A. thaliana* [29•]. These genes are valuable in sorghum because they confer resistance to some insects. The transfer of these sorghum genes into *A. thaliana* resulted in the production of cyanogenic glycosides without causing adverse phenotypic effects in the transgenic plants. Moreover, the transgenic *A. thaliana* plants that expressed the sorghum genes had increased resistance to the flea beetle, a natural pest of other members of the crucifer family.

Conclusions

The surprisingly difficult challenge of identifying and annotating genes in the human genome is being met in part by the acquisition of genome sequences from a range of mammals for comparative analysis. The acquisition of sorghum gene sequences will have similar utility in the annotation of sequences from the genomes of rice, maize, and other grasses. Moreover, the utility of genome analysis in plants is especially clear because plants show tremendous diversity in development, morphology, biochemistry, and their ability to occupy a large number of diverse environmental niches. A comparison of rice, a C3 plant that is well-adapted to wet environments, and sorghum, a C4 plant adapted to arid environments, may elucidate the combination of genetic traits that is required for adaptation to each of these different environments. Rapid progress on the sorghum genome map and EST project, and in the development of other genome technologies, indicates that sorghum will be well positioned for in-depth genome sequence analysis and for comparative analysis of gene function in grasses. We advocate targeted sequencing of the gene-rich portion of the sorghum genome in parallel with the effort to sequence maize. The remainder of the sorghum genome could be sequenced when sequencing costs decline.

Even before the entire sequence is known, genome technologies will soon enable researchers to use sorghum to investigate complex phenomena that have proven inaccessible to in-depth analysis to date. For example, high-resolution AFLP-based genotyping technology and the sorghum genome map will allow researchers to trace precisely the genetic basis of sorghum domestication and selection during more than 50 years of breeding. Similarly, sorghum provides an extraordinary opportunity to investigate the genetic basis of plant adaptation to adverse environments, in which it is a staple in the diet of many of the world's poorest subsistence farmers. The high-resolution sorghum genome map, microarray technology, and advances in our understanding of sorghum gene function will provide the tools and information needed to understand the genetic basis of sorghum's adaptation to adverse environments. A deeper understanding of this fundamental area of plant biology will be the basis of future improvements in crop productivity.

Acknowledgements

Some of the material presented in this review is based upon work supported by the National Science Foundation under Grant No. 0077713. Additional support for research by the authors was derived from the Texas Agricultural Experiment Station, the Perry Adkisson Chair in Plant Biology (JEM), the Advanced Technology Program (000517-0242) (JEM and PEK) and the US Department of Agriculture Agricultural Research Service (USDA-ARS) (RRK).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doggett H: *Sorghum*, edn 2. New York: John Wiley; 1988.

2. Arumuganathan E, Earle ED: Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 1991, 9:208-218.

3. Kong L, Dong J, Hart GE: Characteristics, linkage-map positions, and allelic differentiation of ***Sorghum bicolor*** (L.) Moench DNA simple-sequence repeats (SSRs). *Theor Appl Genet* 2000, 101:438-448.

4. Djè Y, Heuertz M, Lefèbvre C, Vekemans X: Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. *Theor Appl Genet* 2000, 100:918-925.

5. Doebley J, Durbin M, Golenberg EM, Clegg MT, Ma DP: Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (***Gramineae***). *Evolution* 1990, 44:1097-1108.

6. •• Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE: A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Res* 2000, 10:789-807.
A novel method for building genetic and physical genome maps is described that is based on six-dimensional pooling of BAC libraries combined with AFLP technology. The technology is used to construct a framework integrated genetic and physical map of the sorghum genome.

7. Menz MA, Klein RR, Mullet JE, Orbert JA, Unruh NC, Klein PE: A high-density genetic map of ***Sorghum bicolor*** (L.) Moench based on 2926 AFLP, RFLP and SSR markers. *Plant Mol Biol* 2002, in press.

8. •• Childs KL, Klein RR, Klein PE, Morishige DT, Mullet JE: Mapping genes on an integrated sorghum genetic and physical map using cDNA selection technology. *Plant J* 2001, 27:243-255.
Direct selection technology that was originally developed for the human genome project is further developed for high throughput applications in this study. The authors use cDNA direct selection technology to identify genes that are encoded by BACs located on the sorghum genome map.

9. Kim JS, Childs KL, Faridi NI, Menz MA, Klein RR, Klein PE, Price HJ, Mullet JE, Stelly DM: Integrated karyotyping of sorghum by ***in situ*** hybridization of landed BACs. *Genome* 2002, in press.

10. Peng Y, Schertz KF, Cartinhour S, Hart GE: Comparative genome mapping of ***Sorghum bicolor*** (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. *Plant Breed* 1999, 118:225-235.

11. Lin YR, Draye X, Qian X, Ren S, Zhu LH, Tomkins J, Wing RA, Li Z, Paterson AH: Locus-specific contig assembly in highly-duplicated genomes, using the BAC-RF method. *Nucleic Acids Res* 2000, 28:e23. http//www3.oup.co.uk/nar/methods/Volume_28/Issue_07/gnd023_gml.abs.html

12. Draye X, Lin YR, Qian XY, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH: Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. *Plant Physiol* 2001, 125:1325-1341.

13. •• Bennetzen JL: Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* 2000, 12:1021-1029.
A current assessment of the extent and value of conservation of gene order among grass genomes is presented in this timely review. Recent findings concerning the extent of micro-colinearity of genes within grass genomes are discussed relative to the grass genome projects and the goal of understanding grass genome function.

14. Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Liu SC, Burow MD, Kowalski SP, Katsar CS, DelMonte TA *et al.*: Toward a unified genetic map of higher plants, transcending the monocot–dicot divergence. *Nature Genet* 1996, 14:380-382.

15. Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, Düsterhöff A, Stiekema W, Entian KD, Terryn N *et al.*: Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of ***Arabidopsis thaliana***. *Genome Res* 2001, 11:1167-1174.

16. Devos KM, Beales J, Nagamura Y, Sasaki T: ***Arabidopsis***–rice: will colinearity allow gene prediction across the eudicot–monocot divide? *Genome Res* 1999, 9:825-829.

17. Tarchini R, Phyllis B, Robin W, Scott T, Rafalski A: The complete sequence of 340 kbp of DNA around the rice Adh1–Adh2 region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 2000, 12:381-391.

18. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z: Colinearity and its exceptions in orthologous ***adh*** regions of maize and sorghum. *Proc Natl Acad Sci USA* 1999, 96:7409-7414.

19. Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan LL, Shiloff BA, Bennetzen JL: Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol* 2001, 125:1342-1353.

20. Tompa M: Identifying functional elements by comparative DNA sequence analysis. *Genome Res* 2001, 11:1143-1144.

21. Quail PH: The phytochromes: a biochemical mechanism of signaling in sight? *BioEssays* 1997, 19:571-579.

22. Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE: The sorghum photoperiod sensitivity gene, Ma_3, encodes a phytochrome B. *Plant Physiol* 1997, 113:611-619.

23. Finlayson SA, Lee IJ, Morgan PW: Phytochrome B and the regulation of circadian ethylene production in sorghum. *Plant Physiol* 1998, 116:17-25.

24. Finlayson SA, Lee IJ, Mullet JE, Morgan PW: The mechanism of rhythmic ethylene production in sorghum. The role of phytochrome B and simulated shading. *Plant Physiol* 1999, 119:1083-1089.

25. • Zhao ZY, Cai TS, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D: ***Agrobacterium***-mediated sorghum transformation. *Plant Mol Biol* 2000, 44:789-798.
A dramatic increase in *Agrobacterium*-mediated sorghum transformation efficiency is described in this manuscript. Increased transformation efficiency may be due to the use of different selectable markers and the selection of material from field-grown plants for transformation. These changes may help to minimize the gene silencing described by Emani *et al.* [26].

26. Emani C, Sunikumar G, Rathore KS: Transgene silencing and reactivation in sorghum. *Plant Sci* 2002, in press.

27. Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T: Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from ***Sorghum bicolor***. *Proc Natl Acad Sci USA* 1999, 96:15330-15335.

28. Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA *et al.*: Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* 2001, 27:581-590.

29. • Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Høj PB, Møller BL: Resistance to a herbivore through engineered cyanogenic glucoside synthesis. *Science* 2001, 293:1826-1828.
This study demonstrates the feasibility and utility of gene transfer from sorghum, a monocot, to *A. thaliana*, a dicot. This proof of concept paper opens the way for further transfer of valuable genes among a wide range of crop species for the purpose of increasing resistance to pests, stress tolerance, and product quality.